V.B GSSC Database Design

Dave Davis– Data Archive & Software Support

ddavis@milkyway.gsfc.nasa.gov

Database Design Outline

- Requirements

- Data flows between the elements

- Database Design requirements

- Security

- Databases

- Archiving

- Summary

Requirements

- Our requirements are derived from
  - Science Requirements Document (433-SRD-0001)
  - Mission System Specification (433-SPEC-0001)
  - Ground System Requirements Document (433-RQMT-0006)
  - GSSC Functional Requirements Document (433-RQMT-0002)
- Other applicable documents include
  - GLAST Announcement of Opportunity (AO)
  - Project Data Management Plan (PDMP—433-PLAN-0009)
  - Operations Concept Document (433-OPS-0001)
  - GSSC-HEASARC MOU
  - Report of Data Products Working Group

GSSC-Centric Communications / Data Flow
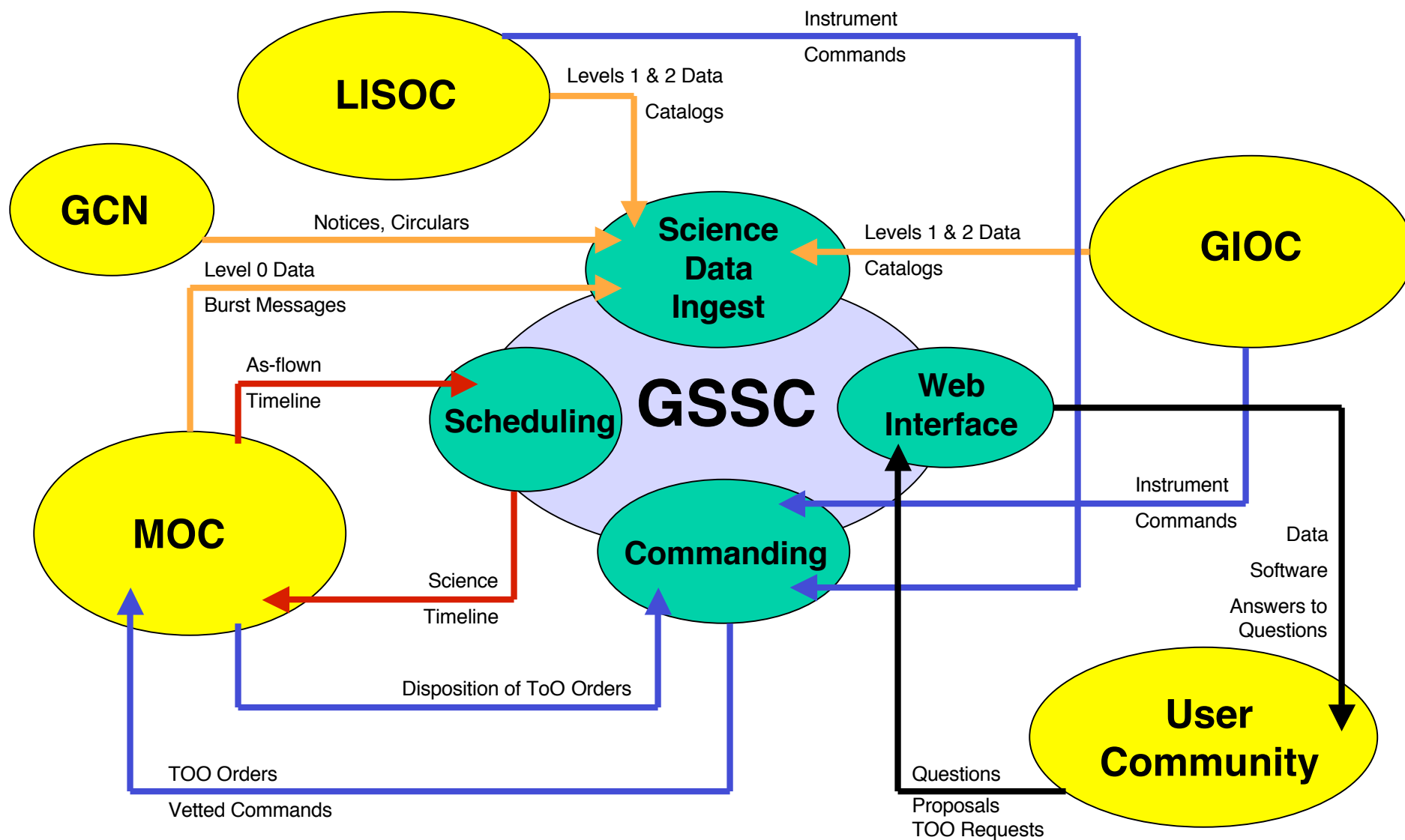
GSSC Databases

- GSSC will maintain the 26 separate databases which are listed in the Design Document

- Three types of databases

  1. Database containing metadata which points to information in files (e.g. HEASARC Browse Database)

  2. Self contained databases (e.g. all data stored and searchable in MySQL database)

  3. Customized data systems (used by GSSC for event and spacecraft pointing data)

- Type 1 in common usage in HEASARC

- Type 2 can be flat files or a DBMS system such as MySQL.

- Type 3 designed with performance, simplicity and reliability as the highest priority design criteria.

Database Design Drivers

- Data products will be available to the public within 24 hrs of the receipt of the data from the MOC and IOC's (FRD 5.6.3).

- The GSSC data products will be HEASARC compatible (FRD 5.7.5).

- All databases

    - must be able to be searched on boolean, time, integers & real values.

    - must be able to store and search at least 10 years worth of mission data.

    - must be remotely accessible for user queries and for data retrieval.

    - must be read-only for external queries.

Database Design Drivers, cont

- Generate GSSC science data products (FRD 5.6.1)

- Produce GSSC science data products at a rate 4x that the data is delivered from the IOC's (FRD 5.6.2)

- GSSC data products must be available within 24 hours of the data receipt from the IOC's and the MOC (FRD 5.6.3)

- Maintain a backup pipeline for the ISOC and the GIOC (FRD 5.6.5)

GSSC System Security

- Security (FRD 5.2.2)
  - System Access will be controlled
    - A firewall exists between the web site and all GSSC computers
    - Password authentication
    - Resource monitoring
  - Archive software will prevent common security "holes"
    - Signal handling
    - Error handling
    - Exception handling

- Security cont.
  - A secure file transfer system (DTS) will be used for file transfer to the HEASARC.
    - DTS is a Leicester developed system to automatically transfer data between sites
      - Currently in use at Goddard for XMM data transfer
      - Swift data will be transferred using DTS
    - The file transfer system will run on a system isolated from the Web
    - Data transfer will be via a secure utility such as sftp

GSSC System Security, cont.

- Data Integrity
  - Utilise a well tested/free DBMS
  - Read-only archive for public access
  - Frequent backups
  - DB agents to check data integrity

- Computer components
  - Standard Linux machines for most applications
  - A Linux cluster where speed is an issue

File Structure

- Primary data products will be available as OGIP compatible FITS files, including the LAT event and photon data.

- Timelines, GCN notices will be available as text files.

- Some products (e.g. the distribution of GBM bursts) will be available as an image on the webpage as well as a FITS catalog.

LAT Databases

A selection of primary LAT databases that are defined and prototyped

ID	Name	Description
D1ev	LAT event data	Merit n-tuple for LAT events
D1ph	LAT photon data	LAT photons selected from the D1ev database
D2	LAT Pointing & Livetime history	LAT Pointing & Livetime history
D3	LAT IRF's	LAT CALDB files
D5	LAT Pt Src	LAT Point Source catalog
D6	LAT Bursts	LAT Burst Catalog
D7	ISM	Interstellar Emission Model
	LAT Transients	LAT Transient Catalog

Note: Orange denotes DB's that are prototyped

GBM Databases

A selection of primary GBM databases that are defined

ID	Name	Description
	GBM CTIME data	GBM data binned in 0.25s time bins and 8 channels
	GBM CSPEC data	GBM data binned in 8.2s time bins and 128 channels
	GBM Burst data	Compilation of GBM data centered on a burst
	GBM Burst Catalog	GBM Burst Catalog
	GBM Burst Spectral Catalog	Catalog of GBM burst spectra

MOC Databases

A selection of primary MOC databases that are defined

ID	Name	Description
	Integrated Observatory Timeline	Timeline with all commands, contacts, etc included
	TRDSS Ephemerides	TRDSS positions
	GLAST Ephemeris	GLAST positional information

Note: These databases may be text files

GSSC Internal Databases

A selection of GSSC tracking databases that are defined

ID	Name	Description
	Data products	Tracks all data products received at the GSSC
	Data processing	Tracks all data products processed at the GSSC
	Process tracking	Tracks all anomalies associated with the processing data products at the GSSC
	Issue Tracking	Internal processing issues
	Data Re-transmission requests	Requests for re-transmission of datasets from the ground elements

Note: These databases are MySQL DB's

Archiving

- All databases will be backed up and archived

- Copies will be kept on hard media

- An archive database will track properties of archival data:

  - Files archived

  - File sizes

  - Checksums

  - All Processing version numbers

  - Dates
    - Received by GSSC
    - Processed by GSSC
    - Archived

  - Physical medium where file is located

  - Minimum expected medium life

Summary

- Database design is structured to fulfill our requirements

- Prototype databases and pipelines show that we are meeting the requirements laid out in the Functional Requirements Document

- System security is integrated into the design and structure of the system components